



PATENT
Atty. Docket No.: 29250-000492/US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application No.: 09/846,200
Filing Date: May 2, 2001
Appellant: Gregory CIURPITA et al.
Group Art Unit: 2665
Examiner: J. Wozniak
Title: METHOD AND APPARATUS FOR AUTOMATIC
RECOGNITION OF LONG SEQUENCES OF
SPOKEN DIGITS
Conf No: 4515

APPELLANTS' BRIEF ON APPEAL UNDER 37 C.F.R. §41.37

U.S. Patent and Trademark Office
Customer Service Window
Randolph Building
401 Dulany Street
Alexandria, VA 22313-1450
Mail Stop Appeal Brief-Patents

March 6, 2006

Dear Sir:

Appellants submit herewith their Brief on Appeal as required by 37 C.F.R.
§41.37.

03/07/2006 SZEWDIE1 00000012 09846200

01 FC:1402

500.00 0P

TABLE OF CONTENTS

	<u>Page</u>
BRIEF ON BEHALF OF APPELLANTS	3
1. REAL PARTY IN INTEREST	3
2. RELATED APPEALS AND INTERFERENCES	3
3. STATUS OF CLAIMS.....	3
4. STATUS OF AMENDMENTS.....	4
5. SUMMARY OF CLAIMED SUBJECT MATTER.....	4
6. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL.....	7
7. ARGUMENTS	7
a. Claims 1-5, 7-8, 10-15, 18, 20-23 and 25-27 are not rendered obvious over Gerson et al. in view of Ammicht et al	7
b. Claims 6, 17 and 19 are not rendered obvious over Gerson in view of Ammicht, and further in view of Hou.....	10
c. Claims 7, 8, 22 and 23 are not rendered obvious over Gerson in view of Ammicht, and further in view of Vanbuskirk.....	11
d. Claims 9 and 24 are not rendered obvious over Gerson in view of Ammicht, and further in view of Ladd	11
e. Claim 16 is not rendered obvious over Gerson in view of Ammicht, and further in view of Ladd	11
8. EVIDENCE AND RELATED APPEALS APPENDICES	12
9. CONCLUSION	12
10. CLAIMS APPENDIX.....	13

BRIEF ON BEHALF OF APPELLANTS

In support of the Notice of Appeal concurrently filed herewith, appealing the Examiner's final rejection mailed October 4, 2005 of each of pending claims 1-27 of the present application which appear in the attached claims appendix (Appendix 10); Appellants hereby provide the following remarks.

(1) **REAL PARTY IN INTEREST:**

The real party in interest is Lucent Technologies, Inc, as evidenced by the assignment recorded at reel 011767, frame 0793.

(2) **RELATED APPEALS AND INTERFERENCES:**

No related appeals or interferences are known.

(3) **STATUS OF CLAIMS:**

Claims 1-27 are pending in the application.

Claims 1-5, 7-8, 10-15, 18, 20-23 and 25-27 remain finally rejected under 35 U.S.C. §103(a) as being unpatentable over Gerson et al. (U.S. Patent No. 4,870,686) in view of Ammicht et al. (U.S. Patent No. 6,246,986).

Claims 6, 17 and 19 remain finally rejected under 35 U.S.C. §103(a) as being unpatentable over Gerson in view of Ammicht, and further in view of Hou et al. (U.S. Patent No. 5,325,421).

Claims 7, 8, 22 and 23 remain finally rejected under 35 U.S.C. §103(a) as being unpatentable over Gerson in view of Ammicht, and further in view of Vanbuskirk et al. (U.S. Patent No. 6,505,155).

Claims 9 and 24 remain finally rejected under 35 U.S.C. §103(a) as being unpatentable over Gerson in view of Ammicht, and further in view of Larson (an article entitled "Investigating a Mixed Initiative Dialogue Management Strategy", 1997).

Claim 16 remains finally rejected under 35 U.S.C. §103(a) as being unpatentable over Gerson in view of Ammicht, and further in view of Ladd et al. (U.S. Patent No. 6,269, 336).

(4) STATUS OF AMENDMENTS:

No amendments have been filed subsequent to the October 4, 2005 Final Office Action.

(5) SUMMARY OF CLAIMED SUBJECT MATTER:

Automatic speech recognition (ASR) or voice recognition (VR) systems have begun to gain widened acceptance in a variety of practical applications. In conventional voice recognition systems, a caller interacts with a voice response unit having a voice recognition capability. Such systems typically either request a verbal input or present the user with a menu of choices, and wait for a verbal response, interpret the response using voice recognition techniques, and carry out the requested action, all typically without human intervention.¹ Further, it has been observed that when the user speaks out sequence of digits, whether short or long (e.g., a telephone number or credit card number), the user tends to do so in natural groups of smaller digit strings or subgroups, such as several digits at a time, with a natural pause between subgroups.

However, conventional systems produce errors in the recognition stage because whenever these conventional voice recognition systems interpret a digit sequence (i.e., a 16-digit credit card number), there is some uncertainty as to the correspondence between the utterance and the interpretation.² Accordingly, in order to deal with these potential errors, conventional systems may use some type of verification for all transactions in situations where the error rate may cause concern, in order to avoid the possibility of processing an incorrect digit string.

For example, following the input of each connected digit string, a voice recognition system may "read back" (i.e., feedback) the best digit string candidate, and require an affirmative or negative response from the individual using the system. An example would be: "please say yes if your credit card number is 1234-5678-9012-3456", and please say "no otherwise". Although this type of verification is often necessary and useful, it is more often cumbersome, time consuming and generally tortuous for frequent users of a voice recognition system.³

¹ Appellants' specification, paragraph [0002].

² Appellants' specification, paragraph [0006].

³ Appellants' specification, paragraph [0005].

In order to overcome the above deficiencies in automatic speech recognition of sequences of spoken speech units, a method and system of recognizing speech in user-interface recognition systems has been disclosed, that is based at least partially on the above observation that a speaker naturally pauses and may speak in smaller subgroups of speech units or digits that form part of a complete longer speech sequence. The system attempts to provide feedback after each subgroup by repeating the recognition results, allowing the user to correct the results if erroneous. Additionally, the method and system take advantage of an observation that a user not only naturally speaks slower when errors in recognition occur, but will also naturally speak in smaller groups of speech units as repeated errors in speech verification occur.⁴

In another exemplary embodiment, an utterance or subgroup of speech units may be received or detected by the system between the aforementioned natural pauses. This pause is detected by the system and the subgroup is processed in order to provide an interpretation or recognition result that is temporarily stored in the system. The recognition result, which is a best representation of the input subgroup, is immediately repeated back to the user for verification. Each recognition result of a subgroup or sequence (i.e., best system interpretation of sequence) is verified by being fed back to the user. For example, if a rejection criteria is met, such as the user rejecting a recognition result by saying "no" for example, the sequence being verified is rejected, and the sequence prior to that (previous result) is fed back for re-verification. The system also provides for multiple occurrences of "no" being uttered by the user, and even mis-recognition of a user's negative utterance by the system itself, by enabling the user to skip back where necessary to correct errors. Otherwise, if there are no errors indicated in the results (such as when the user immediately inputs the next subgroup), the processing steps are repeated for remaining subgroups or sub digit-sequences until it has been determined that the complete speech sequence has been accurately recognized.⁵

In accordance to an exemplary embodiment of the present invention, the system feedback may be embodied as follows:

⁴ Appellants' specification, paragraph [0008].

⁵ Appellants' specification, paragraphs [0009] and [0019].

- (a) the user could listen through the entire feedback (repeated subgroup), and then continue with the next digits-subgroup (i.e., "123" is repeated; the user realizes this is correct and says the next subgroup "456"); or
- (b) the user could hear a mistake in the feedback (incorrect repeated subgroup), so he/she can reject the result (i.e., user hears "457" instead of "456", so he/she says "no" either before or after the feedback completes). In this case the previous subgroup is repeated (e.g., "123") so the user can repeat input of "456"; or
- (c) the user may reject the current results and immediately repeat the subgroup (e.g. "no 456"). In this case, the system will discard the subgroup being rejected, and repeat the recognition results for the speech following "no", without repeating the previous results; or
- (d) the user may also begin speaking the next subgroup without waiting for the repeated results to be completely played back, in which case, the current and previous results are concatenated together and treated as a single subgroup (e.g. "123456"); or
- (e) the user may speak "no" repeatedly, rejecting previously accepted subgroups. This also foresees the scenario where a spoken "no" for a subgroup or series of subgroups was not recognized by the system, contributing to an erroneous result.⁶

Regardless of the number of subgroups or utterances, all recognition results must be confirmed. After being fed back to the user, all recognition results are assumed to be confirmed unless explicitly rejected by the user; such as through a negative command like "no" for example. Moreover, results that have already been confirmed can still be rejected, simply by repeated rejections. Further, even if the initial bad results are not rejected by the user, and subsequent errors are added thereto, the system provides the user the ability to skip back (i.e., to previous, previous-previous subgroup, etc.) where necessary to correct mistakes.⁷

⁶ Appellants' specification, paragraph [0028].

⁷ Appellants' specification, paragraph [0029].

(6) GROUND OF REJECTION TO BE REVIEWED ON APPEAL:

a) Claims 1-5, 7-8, 10-15, 18, 20-23 and 25-27 are rejected under 35 U.S.C. §103(a) as being unpatentable over Gerson et al. (U.S. Patent No. 4,870,686) in view of Ammicht et al. (U.S. Patent No. 6,246,986).

b) Claims 6, 17 and 19 are rejected under 35 U.S.C. §103(a) as being unpatentable over Gerson in view of Ammicht, and further in view of Hou et al. (U.S. Patent No. 5,325,421).

c) Claims 7, 8, 22 and 23 are rejected under 35 U.S.C. §103(a) as being unpatentable over Gerson in view of Ammicht, and further in view of Vanbuskirk et al. (U.S. Patent No. 6,505,155).

d) Claims 9 and 24 are rejected under 35 U.S.C. §103(a) as being unpatentable over Gerson in view of Ammicht, and further in view of Larson (an article entitled "Investigating a Mixed Initiative Dialogue Management Strategy", 1997).

e) Claim 16 is rejected under 35 U.S.C. §103(a) as being unpatentable over Gerson in view of Ammicht, and further in view of Ladd et al. (U.S. Patent No. 6,269,336).

(7) ARGUMENT:

a. Claims 1-5, 7-8, 10-15, 18, 20-23 and 25-27 are not rendered obvious over Gerson et al. in view of Ammicht et al.

Appellants submit that Gerson and Ammicht, individually or in combination, fail to disclose or suggest a method of recognizing speech in systems that accept speech input, comprising, at least:

immediately feeding back the recognition result for verification by the user, wherein the recognition result is interrupted by the user prior to being fed back for verification,

as recited in claim 1.

The Examiner admits that “Gerson does not teach the use of a barge-in feature that allows the user to interrupt a recognition result before it is completely fed back to the user”⁸. Yet, the Examiner attempts to overcome the deficiency in Gerson by arguing that Ammicht discloses the interruption of the recognition result prior to being fed back for verification. Appellants respectfully disagree.

Column 3, lines 34-46, and etc., of Ammicht recites:

FIG. 3 illustrates one implementation of a barge-in arrangement that is designed to help facilitate user-machine interactions in interactive voice response systems (VRUs). The purpose of such an arrangement is twofold. First, the barge-in arrangement allows a machine to ascertain, while it is playing a prompt, whether it is being interrupted by a user uttering meaningful speech, as opposed to simply by noises or language not meant to be heard and used by the machine. Second, the barge-in arrangement distinguishes between a speaker who is merely pausing to think, and a speaker who is done speaking.

That is, Ammicht is directed to a barge-in operation that determines whether the system receives an utterance that is intended to be interrupted or is merely noise. However, the barge-in feature is not performed so as to interrupt the recognition result just before the recognition result is fed back. Ammicht discloses that the barge-in operation is performed during a prompt period, rather than during a recognition result period. Accordingly, Ammicht fails to disclose that the “recognition result is interrupted by the user prior to being fed back for verification”, as recited in claim 1.

Further, the Examiner has failed to provide any evidence or motivation why one skilled in the art would use the interactive voice response unit of Ammicht in the user-interactive speech recognition control system of Gerson. Thus, no evidence has been provided as to why one of ordinary skill in the art would be motivated to incorporate the “barge-in” feature of Ammicht into the system of Gerson, where Gerson discloses that the recognized digits must be completed prior to continuing the digit sequence⁹ (e.g., the recognition result cannot be interrupted by the user prior to being fed back for verification).

⁸ October 4, 2005 Final Office Action dated, page 5, 3rd full paragraph.

⁹ Gerson et al. , col. 3, lines 32-42, col. 7, lines 26-67.

Gerson's shortcomings must be supplemented by some other teaching, and one of ordinary skill in the art must be motivated to provide the supplemental teaching by some motivation, teaching or suggestion of the desirability to make the combination, *In re Dembiczak*, 175 F.3d 994, 50 USPQ2d 1614 (Fed. Cir. 1999), and *In re Kotzab*, 217 F.3d 1365, 55 USPQ 1313 (Fed. Cir. 2000).

Further, the Examiner is using *impermissible hindsight* reconstruction to reject the features recited in claim 1. That is, the Examiner's assertion that:

it would have been obvious to a person of ordinary skill in the art, at the time of the invention, to modify the teachings of Gerson with the "barge-in" feature taught by Ammicht in order to help facilitate user-machine interactions by allowing a user to interrupt a prompt with a meaningful speech input at a time¹⁰

is not evidence for obviousness. Appellants disagree with the Examiner's reasoning, and submit that the mere possibility that one element in one reference could be used in another is not sufficient evidence of a suggestion or motivation to combine the two references.

Appellants submit that the Examiner has used the present application as a blueprint, selected a prior art speech recognition control system of Gerson as the main structure, and then searched other prior art for the missing features (e.g., recognition result is interrupted by the user prior to being fed back for verification), without identifying or discussing any specific evidence of motivation to combine, other than providing conclusory statements regarding the knowledge in the art, motivation and obviousness.

The Federal Circuit has noted that the PTO and the Courts "cannot use hindsight reconstruction to pick and choose among isolated disclosures in the prior art to deprecate the claimed invention", *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1780 (Fed. Cir. 1988), and that the best defense against hindsight based obviousness analysis is the rigorous application of the requirement for a showing of a teaching or motivation to combine the prior art references. Thus, Appellants respectfully submit that the Examiner has failed to provide evidence of motivation for combining the teachings of Gerson and Ammicht.

¹⁰ October 4, 2005 Final Office Action, page 5, 4th full paragraph.

Accordingly, the Examiner has not adequately supported the selection and combination of Gerson and Ammicht to render claim 1 obvious. Claims 2-5 and 10-12, which depend upon claim 1, are also patentable for the reasons stated above with respect to claim 1, as well as on their own merits.

Independent claim 13 is patentable, for reasons similar to those, as discussed above. Specifically, Gerson and Ammicht, individually or in combination, fail to disclose or suggest “a controller for evaluating the output recognition result and feeding back the recognition result to the user, wherein the recognition result is interrupted by the user prior to being fed back for verification”.

Claims 14, 15, 18, 20, 21 and 25-27, which depend upon claim 13, are also patentable for the reasons stated above with respect to claim 13, as well as on their own merits.

Withdrawal of the rejections as to claims 1-5, 7-8, 10-15, 18, 20-23 and 25-27 and allowance of each of these claims is respectfully requested.

b. Claims 6, 17 and 19 are not rendered obvious over Gerson in view of Ammicht, and further in view of Hou

Appellants submit that Gerson, Ammicht and Hou, singly or in any combination, fail to disclose or suggest “a negative utterance representation that is included in the recognition result” and “the rejection criteria is met if the negative utterance is included therein”, as recited in claim 17. The Examiner allegedly asserted that Hou teaches the above features.

Hou is silent as to any teachings of including a “negative utterance” in the recognition result. The Examiner relies on the teachings of Hou for such a teaching. Hou relates to allowing a caller to place a telephone call by merely uttering a label identifying a desired called destination and to charge the telephone call to a particular billing account by merely uttering a label identifying that account.¹¹ Further, Hou teaches that if a subscriber says the word “cancel” after having entered a number of digits, then the CIU 20-1 serving the call in response to detecting that utterance discards the received digits and retransmits the aforementioned announcement.¹² Hou thus fails to teach or suggest the recognition result includes

¹¹ See Hou, Abstract

¹² See Hou, col. 10, lines 42-58.

at least one of a subgroup of speech units and a negative utterance representation that is included in the recognition result, and wherein the rejection criteria is met if the negative utterance is included therein", as recited in claim 17.

Claims 6 and 19 are also allowable for the similar reasons discussed above.

Accordingly, claims 6, 17 and 19 are separately patentable for at least the reason that the Examiner has failed to identify the above feature in any of Gerson, Ammicht and Hou.¹³

c. Claims 7, 8, 22 and 23 are not rendered obvious over Gerson in view of Ammicht, and further in view of Vanbuskirk

As discussed above in section a), Gerson and Ammicht neither disclose nor suggest the claimed invention as found in claims 1 and 13, the independent claims from which the rejected claims depend. Further, Vanbuskirk fails to overcome the noted deficiencies of Gerson and Ammicht. Thus, it is respectfully requested that the rejection be withdrawn.

d. Claims 9 and 24 are not rendered obvious over Gerson in view of Ammicht, and further in view of Larsen

As discussed above in section a), Gerson and Ammicht neither disclose nor suggest the claimed invention as found in claims 1 and 13, the independent claims from which the rejected claims depend. Further, Larson fails to overcome the noted deficiencies of Gerson and Ammicht. Thus, it is respectfully requested that the rejection be withdrawn.

e. Claim 16 is not rendered obvious over Gerson in view of Ammicht, and further in view of Ladd

As discussed above in section a), Gerson and Ammicht neither disclose nor suggest the claimed invention as found in claims 1 and 13, the independent claims from which the rejected claims depend. Further, Larson fails to overcome the noted deficiencies of Gerson and Ammicht. Thus, it is respectfully requested that the rejection be withdrawn.

¹³ Claims 6, 17 and 19 are also allowable for any of the above reasons in section a) by virtue of their dependency from either claims 1 or 13.

(8) EVIDENCE AND RELATED APPEALS APPENDICES:

At there are no related appeals and interferences, copies of a decision rendered by a court or the Board for such proceedings do not exist and have not been supplied in an Appendix pursuant to 41.37(c)(1)(x).

As no evidence was submitted and relied upon in this Appeal, an Appendix pursuant to 37 C.F.R. §41.37(c)(1)(ix) has not been supplied.

(9) CONCLUSION

For all the reasons set forth above, the present invention as recited in Appellants' pending claims 1-27 are not rendered obvious to one skilled in the art as asserted by the Examiner. Accordingly, it is respectfully submitted that the claimed invention should properly be patentable over the cited art. It is therefore respectfully requested that this Appeal be granted by the panel and that the Examiner be reversed.

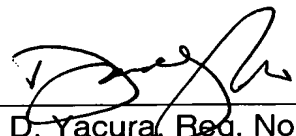
In the event that any matters remain at issue in the application, the Examiner is invited to contact the undersigned at (703) 668-8000 in the Northern Virginia area, for the purpose of a telephonic interview.

If necessary, the Commissioner is hereby authorized in this, concurrent, and future replies, to charge payment or credit any overpayment to Deposit Account No. 08-0750 for any additional fees required under 37 C.F.R. § 1.16 or under 37 C.F.R. § 1.17; particularly, extension of time fees.

Respectfully submitted,

HARNESS, DICKEY, & PIERCE, P.L.C.

By


for Gary D. Yacura, Reg. No. 35,416
P.O. Box 8910
Reston, Virginia 20195
(703) 668-8000

GDY/MJL/DJC

Attached: (10) Appendix: Pending claims of record

(10) CLAIMS APPENDIX:

1. A method of recognizing speech in systems that accept speech input, comprising:

(a) receiving at least a current subgroup of speech units that form part of a complete speech sequence that is to be input from a user, the complete speech sequence being embodied as at least one of a word and a password comprised of a plurality of alphanumeric characters, the subgroup being one or more alphanumeric characters of the complete speech sequence;

(b) detecting a natural pause between input subgroups such that a pause between two alphanumeric characters in a given subgroup or a pause between one alphanumeric character and a subgroup are detected;

(c) recognizing the speech units of the subgroup to provide a recognition result; and

(d) immediately feeding back the recognition result for verification by the user, wherein the recognition result is interrupted by the user prior to being fed back for verification.

2. The method of claim 1, wherein said user is only prompted to repeat said subgroup for re-recognition and re-verification if a rejection criteria is met.

3. The method of claim 1, further comprising:

(e) repeating steps (a) to (d) for remaining input until it is determined that the complete speech sequence has been recognized.

4. The method of claim 1, wherein step (d) is effected using pre-recorded prompts or via text-to-speech synthesis, (TTS) to feedback the recognition result.

5. The method of claim 2, wherein said rejection criteria is embodied as a negative utterance spoken by the user after receiving the fed back recognition result.

6. The method of claim 2, wherein said rejection criteria is embodied as a negative utterance spoken by the user concurrent with inputting the subgroup that is recognized in step (c).

7. The method of claim 2, wherein if said rejection criteria are met repeatedly, the user is prompted to speak the subgroups in smaller groups of speech units.

8. The method of claim 7, wherein said prompt to speak subgroups in smaller groups of speech units provides a built in training mechanism for the user.

9. The method of claim 2, wherein if said rejection criteria are met repeatedly, the user is prompted to use a dial pad to enter the speech units.

10. The method of claim 1, wherein said speech units are selected from any of spoken digits and spoken letters.

11. The method of claim 1, wherein input of a next subgroup after receiving the fed back recognition result indicates a correct recognition of the currently input subgroup.

12. The method of claim 2, wherein said rejection criteria requires determining a level of confidence in said recognition result.

13. An automatic speech recognition system, comprising:
a receiver for receiving at least a current subgroup of speech units that form part of a complete speech sequence that is to be input by a user, the complete speech sequence being embodied as at least one of a word and a password comprised of a plurality of alphanumeric characters, the subgroup being one or more alphanumeric characters of the complete speech sequence;
a detector for detecting a natural pause after receiving the alphanumeric characters;

a speech recognition unit for detecting a natural pause between input subgroups to output a recognition result representative of the current subgroup such that a pause between two alphanumeric characters in a given subgroup or a pause between one alphanumeric character and a subgroup are detected; and

a controller for evaluating the output recognition result and feeding back the recognition result to the user, wherein the recognition result is interrupted by the user prior to being fed back for verification.

14. The system of claim 13, wherein said user is only prompted to repeat said subgroup for re-recognition and re-verification if a rejection criteria is met.

15. The system of claim 13, wherein the speech recognition unit compares the input subgroup with stored recognition grammar in order to determine the recognition result.

16. The system of claim 15, wherein the recognition grammar is stored in a remote memory accessible by the speech recognition unit.

17. The system of claim 14,
wherein the recognition result includes at least one of a subgroup of speech units and a negative utterance representation that is included in the recognition result, and

wherein the rejection criteria is met if the negative utterance is included therein.

18. The system of claim 14, wherein said rejection criteria is met if the user speaks a negative utterance after receiving the fed back recognition result.

19. The system of claim 14, wherein said rejection criteria is met if the user speaks a negative utterance while inputting the current subgroup, so that said recognition result includes the negative utterance.

20. The system of claim 14, wherein the system remains active to process subsequent subgroups until it is determined that the complete speech sequence has been recognized.

21. The system of claim 13, wherein said controller accesses pre-recorded prompts or a text-to-speech synthesis processor in order to effect feedback of the recognition result to the user.

22. The system of claim 14, wherein if said rejection criteria is met repeatedly, said controller prompts the user to speak the subgroups in smaller groups of speech units.

23. The system of claim 22, wherein said prompt to speak subgroups in smaller groups of speech units provides a built in training mechanism for the user.

24. The system of claim 14, wherein if said rejection criteria is met repeatedly, said prompt generator prompts the user to use a dial pad to enter digits corresponding to the speech units.

25. The system of claim 13, wherein said speech units are selected from any of spoken digits, spoken letter and spoken words.

26. The system of claim 13, wherein input of a next subgroup after receiving the fed back recognition result indicates a correct recognition of the currently input subgroup.

27. The system of claim 13, wherein said speech recognition unit determines a confidence level for said recognition result.